

Сравнительный анализ систем машинного перевода

Южанинов Руслан Александрович

Студент (бакалавр)

Алтайский государственный педагогический университет, Лингвистический институт,
Барнаул, Россия

E-mail: ruslanking245@gmail.com

Настоящее исследование посвящено сравнительному анализу систем машинного перевода. Оценка результатов тестирования четырёх систем машинного перевода (Google Translate, Яндекс.Переводчик, PROMT, Microsoft Translator) выполнялась на материале технической документации, а также текстов разговорного стиля, содержащих языковые единицы, затрудняющие автоматический перевод.

Как известно, одним из инструментов переводчика, повышающих эффективность работы и сокращающих временные и трудовые затраты, является машинный перевод. Машинный перевод – это перевод с одного естественного языка на другой, выполняемый компьютерной программой [1]. На сегодняшний день наблюдается расширение сфер применения систем машинного перевода. При помощи данных инструментов осуществляется перевод как устных, так и письменных текстов.

Актуальность исследования обусловлена необходимостью выбора наиболее надёжного инструмента автоматического перевода для работы с текстами разных типов. Фактическим материалом исследования послужил корпус примеров из текстов различной жанрово-стилистической направленности, включающий такие языковые единицы, как термины, фразеологизмы, сленговые выражения, эвфемизмы, явления синтаксической неоднозначности, а также другие синтаксические конструкции. Отбор осуществлялся на основе критерия лингвистической сложности: каждый случай содержал элементы, потенциально вызывающие затруднения у систем автоматического перевода. Фрагмент технической документации был выбран для оценки точности передачи терминологии и соблюдения соответствующих жанрово-стилистических норм. Остальные примеры, иллюстрирующие переводческие соответствия, отобранные машинными переводчиками, были заимствованы из текстов разговорного стиля. Каждый пример переводился четырьмя системами: Google Translate, Яндекс.Переводчик, PROMT и Microsoft Translator. Оценка качества производилась по следующим критериям: 1. точность передачи смысла (сохранение ключевой информации); 2. соответствие нормам русского языка (лексическим, синтаксическим); 3. достижение стилистической адекватности (при передаче, например, сленга, эвфемизмов, регистра); 4. точность терминологии (для технического текста). Каждый вариант перевода оценивался бинарно (0 – не выполнено, 1 – выполнено), затем баллы суммировались. Максимальное количество баллов, которые можно было набрать за все примеры – 9. Данный подход позволил наглядно сравнить системы по числу успешно решённых лингвистически сложных задач. Приведем в качестве примера следующие переводческие соответствия, позволяющие судить об оценке качества перевода систем машинного перевода.

1. Идиомы: *to cost an arm and a leg* – обойтись в целое состояние (Google Translate), стоить руки и ноги (Я.Переводчик), стоить руки и ноги (PROMT), стоить целое состояние (Microsoft Translator).

2. Эвфемизмы: *to cross the rainbow bridge* – уйти в мир иной (Google Translate), перейти радужный мост (Я. Переводчик), пересечь радужный мост (PROMT), уйти через радужный мост (Microsoft Translator).

3. Сленг: *just so much applesauce* – просто полнейшая чушь (Google Translate), просто яблочное пюре (Я. Переводчик), просто такое яблочное (PROMT), просто пустые слова (Microsoft Translator).

4. Предложения с эффектом садовой дорожки: *The old man the boats* – Старик и лодки (Google Translate); Старик с лодками (Я. Переводчик); Старик лодки (PROMT); Старики управляют ложками (Microsoft Translator); *The farmer threw the cow over the fence some hay* – Фермер перебросил корову через забор, оставив немного сена (Google Translate); Фермер перекинул короче через забор немного сена (Я. Переводчик); Фермер бросил корову через забор немного сена (PROMT); Фермер бросил корове через забор немного сена (Microsoft Translator).

5. Термины: фазный провод – *live wire* (Google Translate), *phase wire* (Я. Переводчик), *phase wire* (PROMT), *live wire* (Microsoft Translator); ток срабатывания – *trip current* (Google Translate), *trip current* (Я. Переводчик), *response current* (PROMT), *tripping current* (Microsoft Translator).

6. Синтаксические конструкции: при помощи двухполюсной розетки – *via a two-pole socket* (Google Translate), *using a two-pole socket* (Я. Переводчик), *using a two-pole outlet* (PROMT), *using a two-pole socket* (Microsoft Translator); имеющий время срабатывания – *with a tripping time* (Google Translate), *having a trip time* (Я. Переводчик), *having a response time* (PROMT), *with a tripping time* (Microsoft Translator)

В ходе анализа выявлено, что Google Translate демонстрирует наилучшие результаты в распознавании идиоматических выражений (*to cost an arm and a leg*), сленга (*beans*, *bird course*), эвфемизмов (*to cross the rainbow bridge*), а также при передаче терминологии и синтаксической структуры технических текстов. Microsoft Translator занимает вторую позицию, уступая лидеру в стилистической адаптации и краткости формулировок (*network protection circuit breaker* вместо *circuit breaker*). Яндекс.Переводчик и PROMT показали недостаточную эффективность при работе с нестандартными синтаксическими конструкциями (*The old man the boats* – эффект садовой дорожки) и сленгом (*bird course*, *applesauce*). Таким образом, Google Translate признаётся наиболее предпочтительной системой для решения задач, требующих высокой точности перевода, включая обработку текстов как с когнитивной, так и эмоциональной информацией.

Результаты исследования демонстрируют, что даже современные нейросетевые и гибридные системы машинного перевода сохраняют уязвимость перед явлениями, требующими глубокого семантико-прагматического анализа. Это подтверждает необходимость дальнейшего совершенствования алгоритмов в области обработки идиом, эвфемизмов и явлений синтаксической неоднозначности. Практическая значимость работы заключается в возможности использования полученных результатов для обоснованного выбора системы машинного перевода при решении конкретных переводческих задач. Дальнейшие исследования могут быть направлены на расширение корпуса тестируемых текстов.

Источники и литература

- 1) Семенов А.Л. Современные информационные технологии и перевод : учеб. пособие для студ. перевод. фак. высш. учеб. заведений. М. : Издательский центр «Академия», 2008.